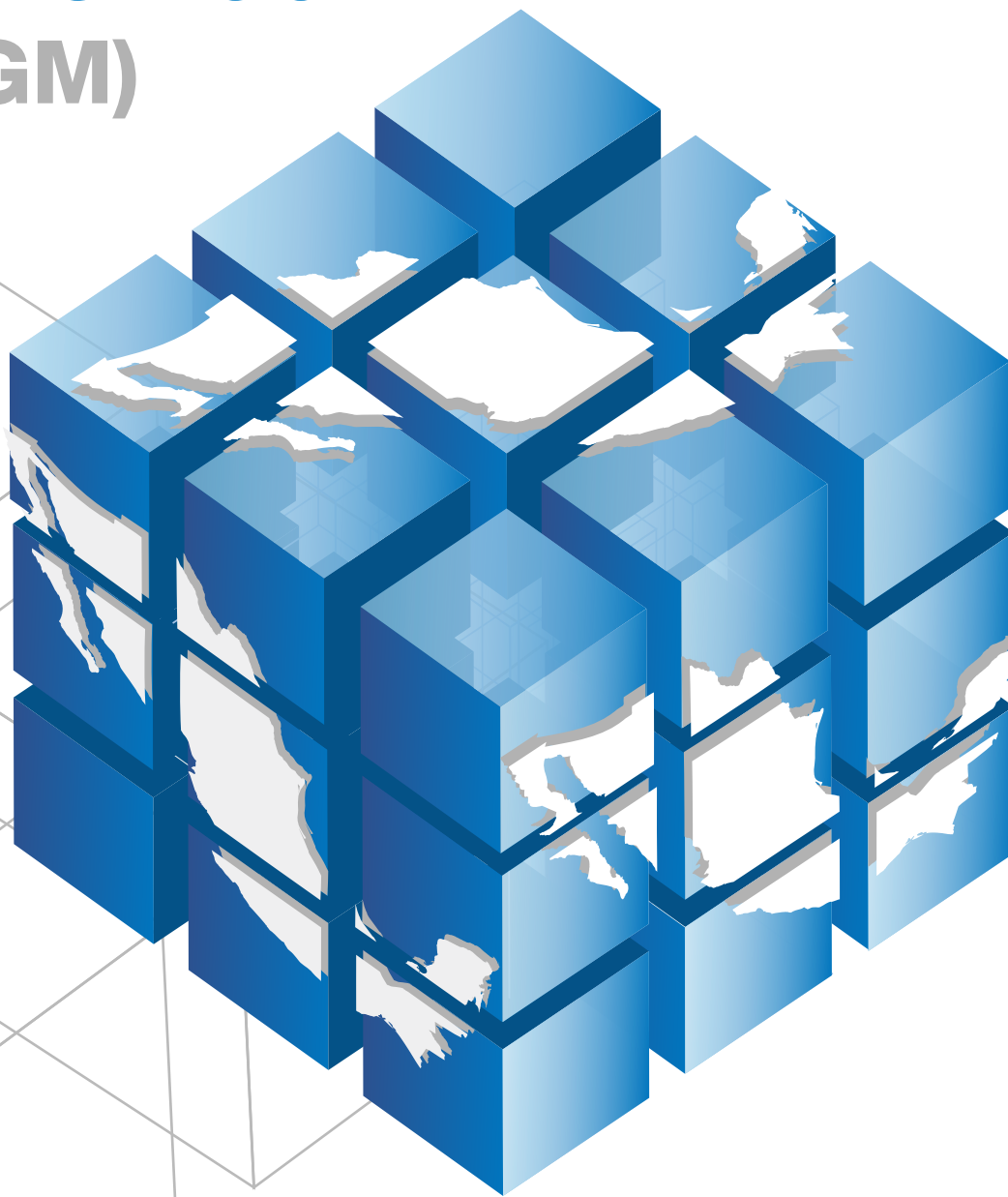
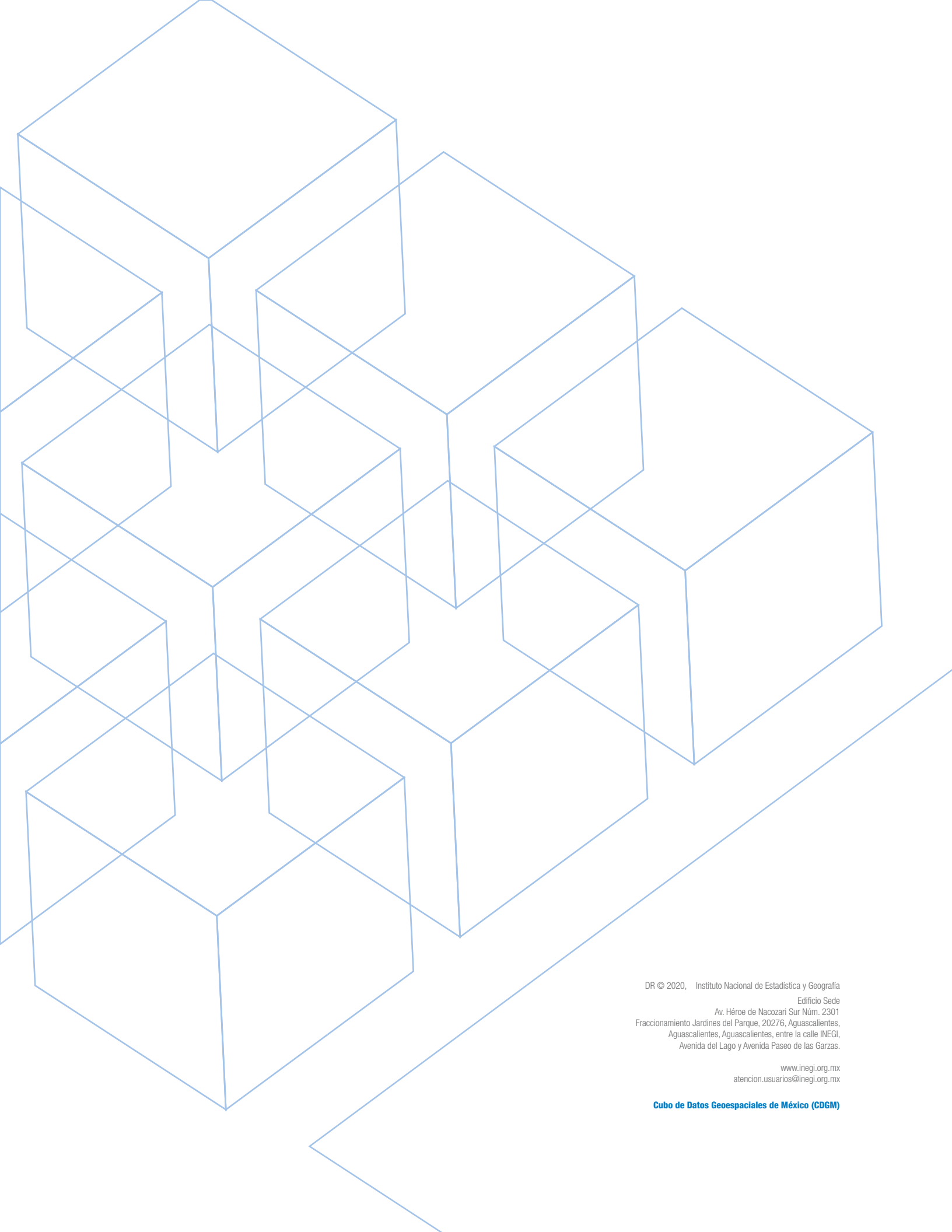


Cubo de Datos Geoespaciales de México (CDGM)





DR © 2020, Instituto Nacional de Estadística y Geografía
Edificio Sede
Av. Héroe de Nacozari Sur Núm. 2301
Fraccionamiento Jardines del Parque, 20276, Aguascalientes,
Aguascalientes, Aguascalientes, entre la calle INEGI,
Avenida del Lago y Avenida Paseo de las Garzas.

www.inegi.org.mx
atencion.usuarios@inegi.org.mx

Cubo de Datos Geospaciales de México (CDGM)

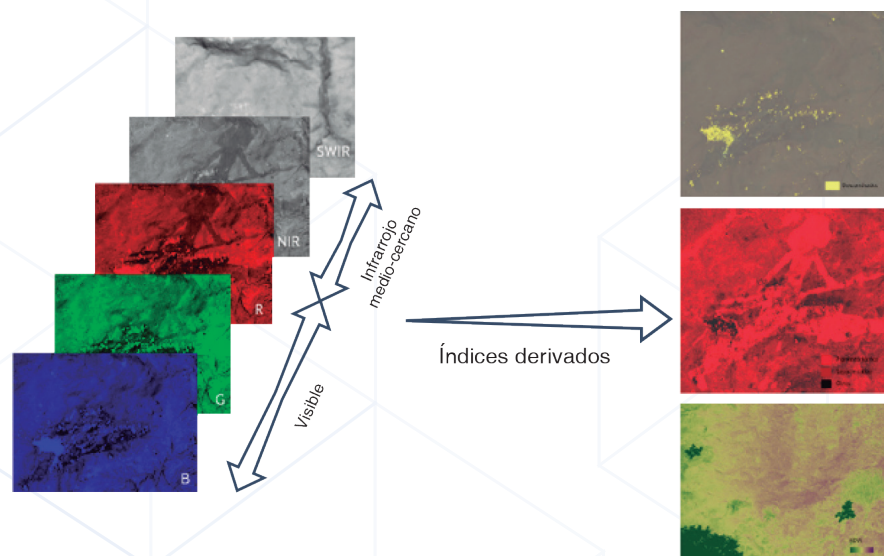
Cubo de Datos Geoespaciales de México (CDGM)

El INEGI tiene la tarea de monitorear fenómenos ambientales, socioeconómicos y demográficos que se presentan en todo el país; es la institución encargada de la cartografía nacional y ha producido mapas de México durante varias décadas. El artículo 26 de la *Ley del Sistema Nacional de Información Estadística y Geográfica* indica que: “El Subsistema Nacional de Información Geográfica y del Medio Ambiente, en su componente geográfico, generará como mínimo los siguientes grupos de datos: marco de referencia geodésico; límites costeros, internacionales, de las entidades federativas, municipales y las demarcaciones territoriales de la Ciudad de México; datos de relieve continental, insular y submarino; datos catastrales, topográficos, de recursos naturales y clima, así como nombres geográficos...”.

Con el CDGM se busca extraer el máximo potencial de las imágenes satelitales y se estableció con la finalidad de contribuir en las tareas encomendadas al INEGI relativas al monitoreo de los datos mencionados.

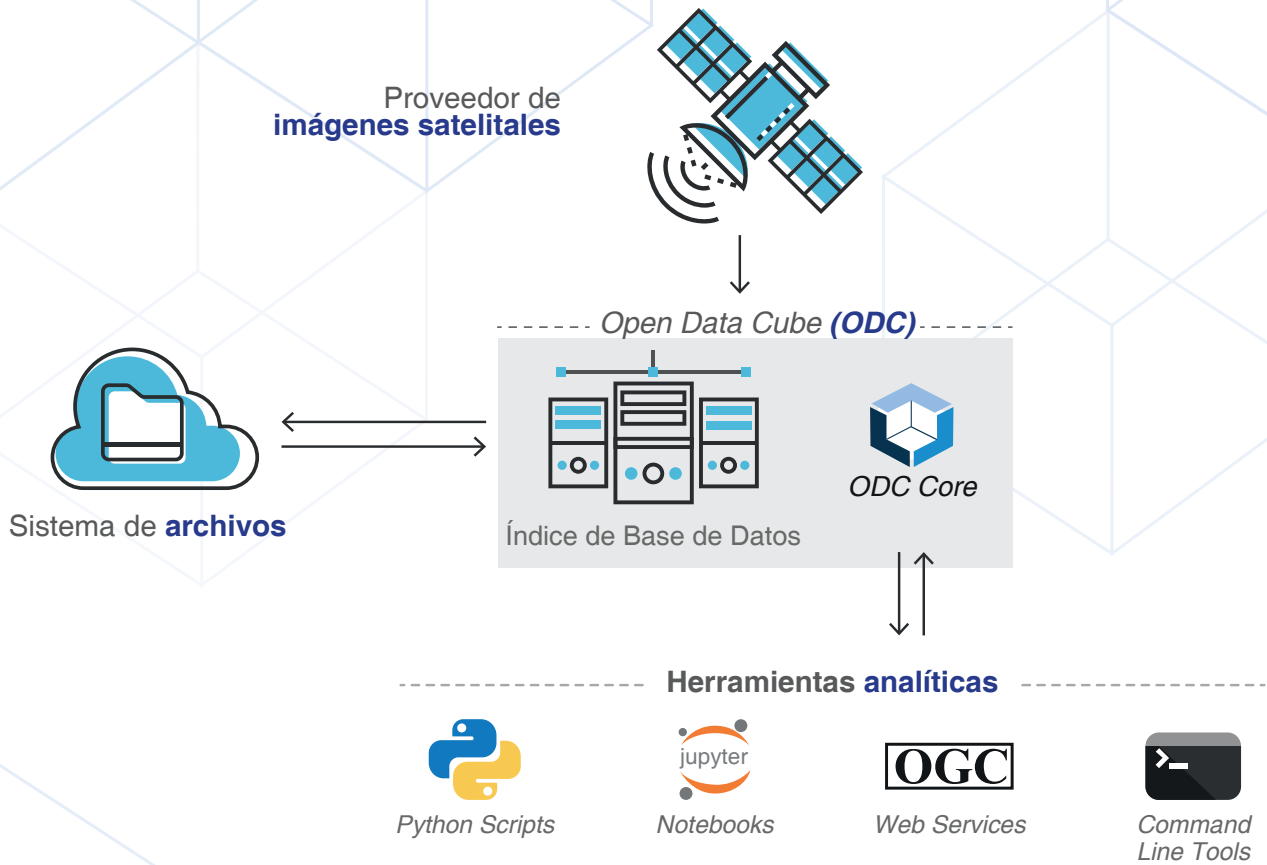
¿Qué es el **CDGM**?

La plataforma tecnológica del CDGM permite acceder, procesar y analizar grandes volúmenes de imágenes satelitales. Estas representan un insumo ideal para la generación de información debido a su creciente disponibilidad, objetividad y periodicidad.

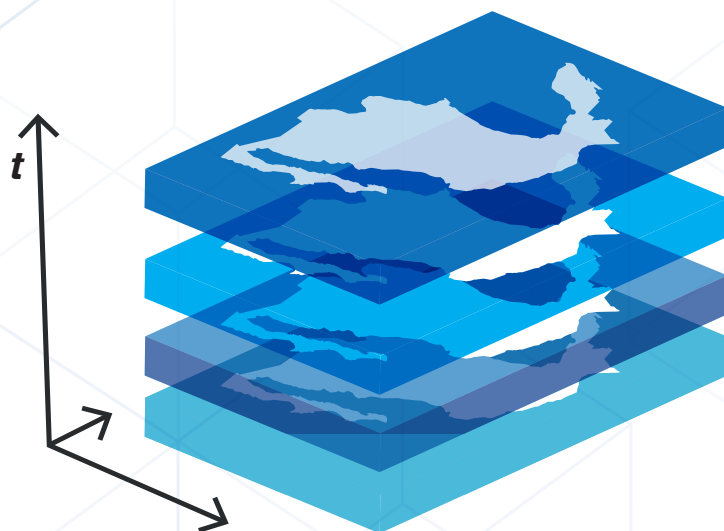


Sin embargo, el aprovechamiento de estos grandes volúmenes de imágenes, que se representan como datos altamente estructurados, conlleva retos tecnológicos y capacidades técnicas especializadas.

El CDGM es una herramienta útil para confrontar estos retos: por un lado, la plataforma del CDGM garantiza un manejo inteligente de los datos satelitales que optimiza los recursos y capacidades computacionales; por el otro, con su interfaz el usuario final puede recuperar imágenes en un formato listo para su análisis. Estas capacidades permiten aprovechar las imágenes satelitales para la generación de información de manera más oportuna.



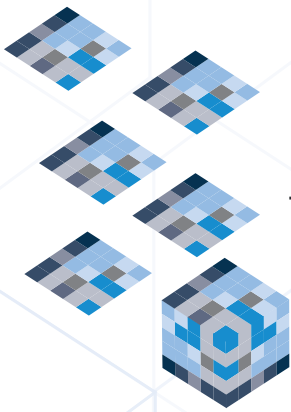
El Cubo gestiona los grandes volúmenes de imágenes aprovechando arreglos masivos de datos tipo *raster*, con los cuales es posible ordenar las imágenes como datos espacio-temporales, lo que facilita su procesamiento y análisis.



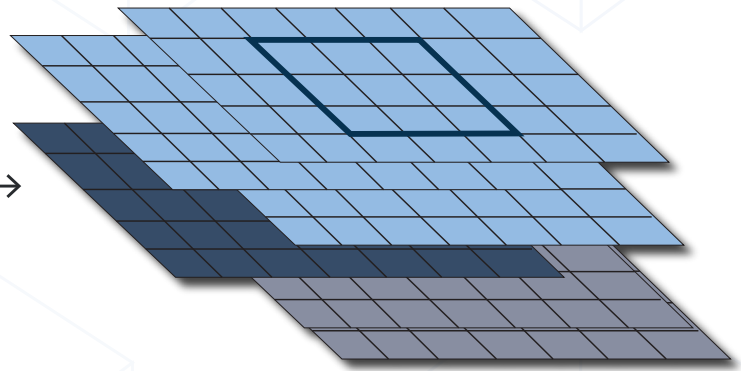
Arreglo masivo de datos multidimensionales (*raster*)

Fuente: INEGI.

De esta manera, todas las imágenes del CDGM son una sola estructura multidimensional, descrita por varios ejes; las coordenadas en estos son el mecanismo que permite acceder a los datos de forma inequívoca; este sistema de coordenadas es homologado para todos los datos, y la ubicación de los píxeles se vuelve independiente de la imagen que lo generó.



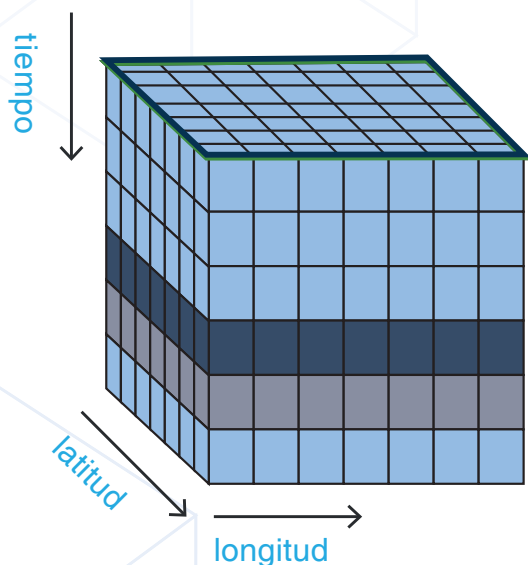
Archivos de imágenes
TIF, netCDF, etcétera



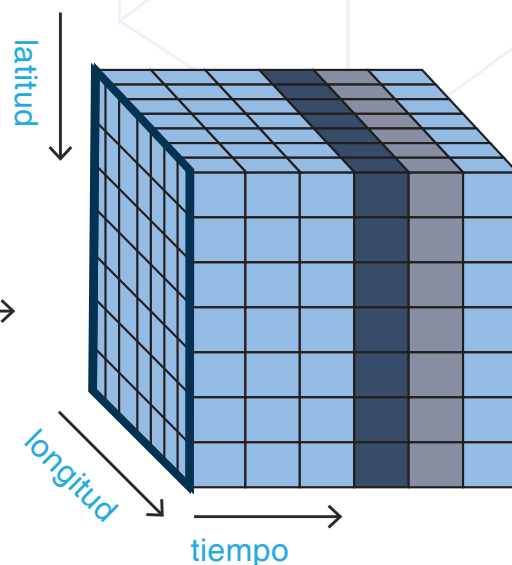
Conjunto de mosaicos
Cubo de imágenes

Open Data Cube (ODC)

Es el sistema informático (<https://www.opendatacube.org/>) que soporta la operación del CDGM. Esta tecnología proporciona una herramienta de explotación de código abierto y de libre acceso, además de que fomenta una comunidad para desarrollar, mantener y aumentar la amplitud y profundidad de las aplicaciones.



Alineamiento de pixeles
Cubo de imágenes



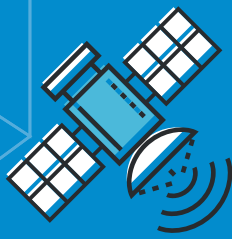
Transpuesto
Cubo de imágenes

Técnicamente, el CDGM es un conjunto de bibliotecas *Python*, una base de datos *PostgreSQL* y un acervo de imágenes *Landsat* ordenadas. Estos recursos se encuentran alojados en los servidores del INEGI.

Las funciones de las bibliotecas permiten catalogar, indexar y procesar los miles de imágenes ordenadas del acervo a partir del formato que tienen cuando se descargan del Servicio Geológico de los Estados Unidos (USGS, por sus siglas en inglés). La base de datos *PostgreSQL* del CDGM no guarda las imágenes, sino la ruta donde se ubica cada una dentro del almacenamiento al que tienen acceso los servidores. Además, registra los metadatos de cada imagen, los cuales contienen información acerca de la misma (por ejemplo: su

Ecosistema de la herramienta ODC

Software para el manejo y análisis de datos geoespaciales



Datos satelitales

Ejemplos:

- Landsat.
- Sentinel.
- MODIS.



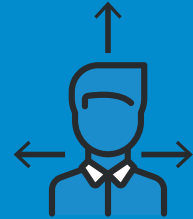
Núcleo



Algoritmos



Aplicaciones



Decisiones informadas

Ejemplos:

- Deforestación.
- Calidad del agua.
- Minería ilegal.

Desarrollo flexible

Dependiendo de tu aplicación el *Open Data Cube* puede ser desplegado en plataformas de cómputo de alto rendimiento, infraestructura en la nube o instalaciones locales. Las instalaciones normalmente corren en sistemas *Linux*, *MacOS* y *Windows*.

El *Open Data Cube* es una solución de código abierto para el acceso, manejo y análisis de grandes cantidades de datos de sistemas de información geográfica, principalmente datos de observaciones de la Tierra.

Fuente: OpenDataCube.org

fuentes, la fecha en la que fue capturada, la región geográfica a la que corresponde y las bandas espectrales que contiene, entre otros datos) que son relevantes a la hora de generar productos derivados; así, el CDGM es un medio que facilita el manejo y acceso a grandes volúmenes de imágenes satelitales a través de una interfaz de programación (basada en el lenguaje *Python*).

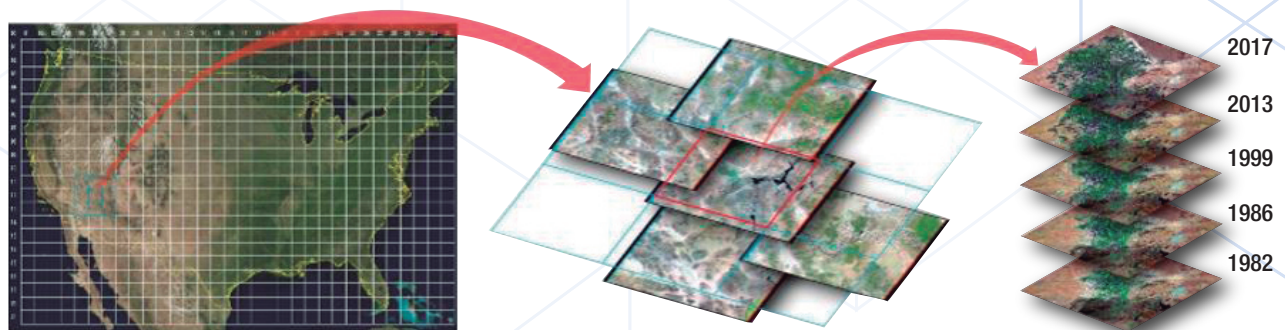
Desde su inicio a principios de 2017, el *software ODC* ha logrado un progreso significativo en el avance de las herramientas y algoritmos de código abierto que admiten la implementación y operación de cubos de datos en sistemas informáticos locales o en la nube en todo el mundo.

La tecnología *ODC* incluye algoritmos que facilitan diversas aplicaciones (como la detección de cambios, la clasificación de cobertura terrestre y la detección de agua mediante aprendizaje computacional).

Imágenes *Landsat* *ARD*

Las imágenes satelitales ideales para la construcción de esta herramienta requieren ser preprocesadas hasta lograr un alto nivel de calidad y estandarización. A esto se le conoce como *datos listos para el análisis* (ARD, por sus siglas en inglés), en los que el valor de cada pixel representa el mismo espacio y muestra la misma respuesta espectral independientemente del sensor, del lugar y de la época en las que fueron captadas, estandarizando, así, las imágenes para su uso estadístico y analítico a través de series de tiempo.

Este nivel de calidad ARD se aplica para varios sensores. En el contexto específico de la misión *Landsat* se conoce como Nivel 1 (T1) y corresponde a escenas calibradas y consistentes con un máximo de 12 metros de error cuadrático medio (RMSE). Adicionalmente, hay otros dos niveles de calidad disponibles para estos sensores: el 2 (T2), que se asigna a escenas calibradas pero que no alcanzan los



Fuente: U. S. Landsat Analysis Ready Data (<https://www.usgs.gov/>).

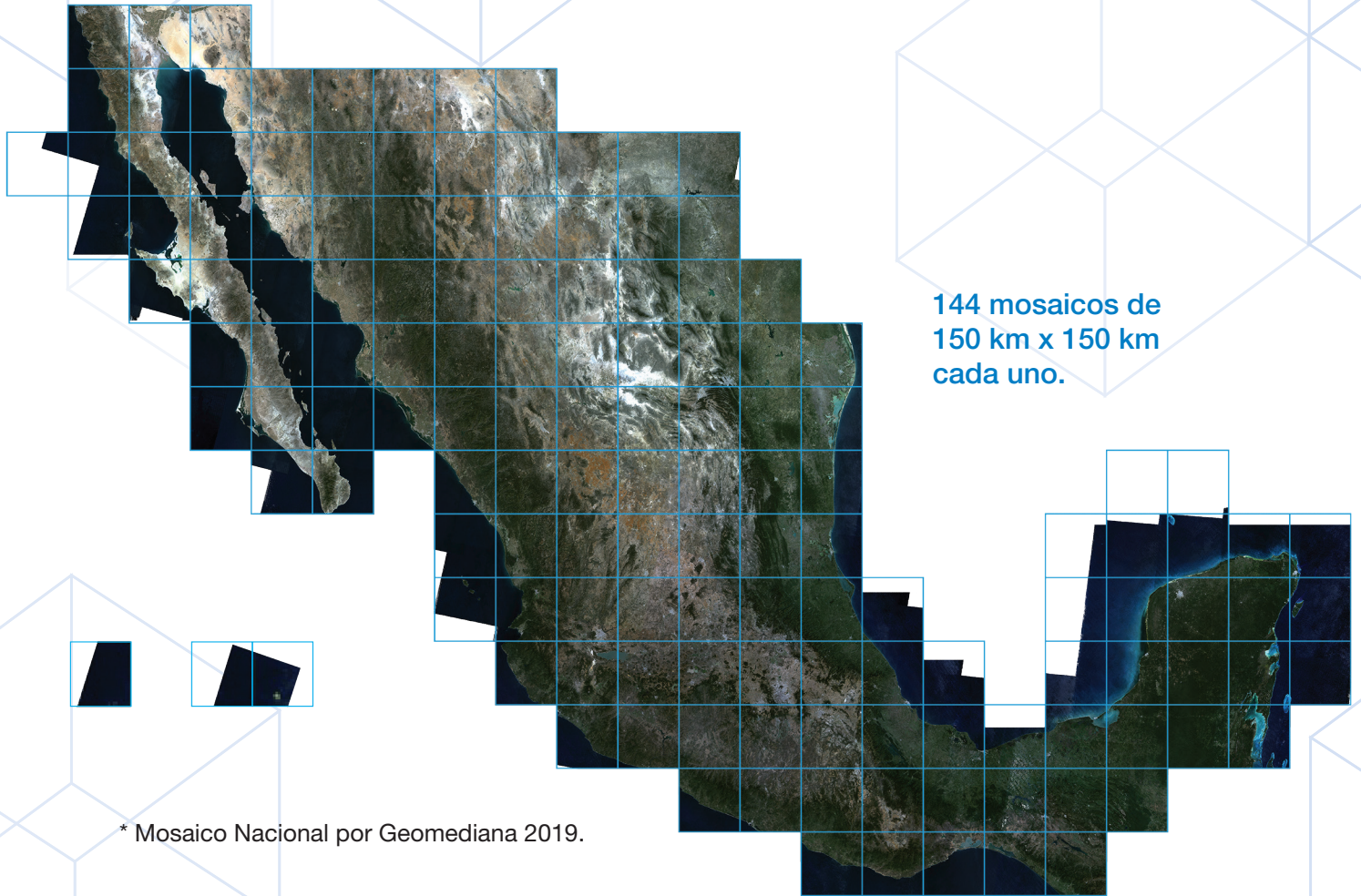
estándares del 1, ya que tienen un RMSE mayor a 12 metros; y, por otro lado, las imágenes Nivel Tiempo Real (RT), que son escenas con procesamiento preliminar y sin ningún proceso de calibración (para información a detalle, consultar la Colección 1 de la Administración Nacional de Aeronáutica y el Espacio, NASA por sus siglas en inglés, y sus niveles de calidad: <https://www.usgs.gov/land-resources/nli/landsat/landsat-collection-1>).

En cuanto a la calidad de las imágenes dentro del CDGM, el acervo corresponde en gran proporción a imágenes ARD: 86.4 % son del Nivel 1 de Calidad de la Colección 1 de la NASA; 13.2 %, del 2; y 0.4 %, del RT.

Aprovechamiento de las **capacidades** del **CDGM**

Generación de mosaicos de compuestos

El CDGM facilita la generación de estos, los cuales aprovechan la información de todas las imágenes de un periodo específico, tal es el caso del mosaico de mediana geométrica o geomediana. Este es de gran utilidad para realizar un resumen estadístico multivariado de to-



144 mosaicos de
150 km x 150 km
cada uno.

* Mosaico Nacional por Geomediana 2019.

dos los valores observados de un mismo pixel en un lapso indicado. El CDGM permite la implementación de un algoritmo que produce una imagen compuesta de *pixeles resumen* de una región específica en un tiempo determinado.

El caso de la geomediana es de especial utilidad, pues mantiene consistencia espacial, incluso, en los límites entre escenas; típicamente, se obtienen imágenes continuas (y libres de nubes). Además, ya que el algoritmo trabaja con todas las bandas del pixel a la vez, se conserva la proporción entre estos valores, por lo que es posible utilizarla como el insumo de otros procesos de uso frecuente que parten de operaciones con las bandas de un pixel; por ejemplo, índices de vegetación, entre otros.

Integración con información estadística

Uno de los grandes desafíos en la actualidad es el reconocimiento de las interconexiones e interdependencias entre los fenómenos sociales, físicos y naturales que suceden sobre la superficie terrestre. El aprovechamiento de las posibilidades descriptivas y predictivas del CDGM, mediante su integración con la información estadística oficial generada por el INEGI, permitirá vincular ambos tipos de datos en análisis orientados a identificar crecimientos, comportamientos, distribuciones, tendencias, patrones y relaciones entre las diversas variables.

Con ello en mente, sería posible, entonces, conocer y reconocer el carácter geoespacial, temporal y sistémico de los fenómenos implicados en temáticas como: crecimiento poblacional, deforestación, contaminación, estrés hídrico, salud, vegetación, protección civil, alimentación, gentrificación, energías limpias, niveles socioeconómicos, movilidad, indigenismo, etcétera.

Aplicaciones de aprendizaje computacional

El interés en el análisis estadístico de los datos satelitales para producir mediciones del medio ambiente, la agricultura y el desarrollo sostenible se establece y continúa aumentando; esto está llevando a una interacción creciente entre la ciencia de la Tierra y los dominios estadísticos, como el aprendizaje automático o *Machine Learning*.

Un actual enfoque clave del análisis con imágenes satelitales en la comunidad de investigación es derivar estadísticas ambientales y agrícolas: el cambio de la cobertura del suelo, la identificación de cultivos, la deforestación y la calidad del agua son algunos ejemplos que hoy se derivan del análisis de datos satelitales. La Organiza-

ción de las Naciones Unidas promueve y fomenta la exploración y el uso de estos datos para monitorear y apoyar la implementación de los Objetivos de Desarrollo Sostenible de la *Agenda 2030 para el Desarrollo Sostenible*. Existe, además, un creciente interés internacional sobre el uso de imágenes satelitales y otras fuentes de *Big Data* para examinar la relación entre los indicadores ambientales y socioeconómicos, como la pobreza.

Este interés en medir los recursos naturales y las estadísticas oficiales a partir de fuentes de *Big Data* disponibles gratuitamente ha llevado a iniciar una vinculación entre las disciplinas estadísticas y de teledetección; como resultado de ello, se están desarrollando aplicaciones de algoritmos de aprendizaje computacional para responder preguntas relacionadas con estos temas.

Referencias

Killough, B. “Overview of the Open Data Cube Initiative”, en: *IGARSS 2018: IEEE International Geoscience and Remote Sensing Symposium*. 2018 (DE) <https://doi.org/10.1109/igarss.2018.8517694>

Roberts, D., B. Dunn & N. Mueller. “Open Data Cube Products Using High-Dimensional Statistics of Time Series”, en: *IGARSS 2018: IEEE International Geoscience and Remote Sensing Symposium*. 2018, July (DE) <https://doi.org/10.1109/igarss.2018.8518312>

Lewis, A., S. Oliver, L. Lymburner, B. Evans, L. Wyborn, N. Mueller & L.-W. Wang (2017). “The Australian Geoscience Data Cube-Foundations and lessons learned”, en: *Remote Sensing of Environment*. 202, 2017, pp. 276-292 (DE) <https://doi.org/10.1016/j.rse.2017.03.015>

Kopp, S., P. Becker, A. Doshi, D. J. Wright, K. Zhang & H. Xu. “Achieving the Full Vision of Earth Observation Data Cubes”, en: *Data*. 4(3), 2019, p. 94 (DE) <https://doi.org/10.3390/data4030094>

Holloway, J. & K. Mengersen. “Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review”, en: *Remote Sensing*. 10(9), 2018, p. 1365 (DE) <https://doi.org/10.3390/rs10091365>

Conociendo México

800 111 46 34

www.inegi.org.mx

atencion.usuarios@inegi.org.mx

